

## Analyzing the Interaction Between Facilitator and Participants in Two Variants of the Think-aloud Method

Maaïke J. van den Haak  
University of Twente  
[m.j.vandenhaak@utwente.nl](mailto:m.j.vandenhaak@utwente.nl)

Menno D.T. de Jong  
University of Twente  
[m.d.t.dejong@utwente.nl](mailto:m.d.t.dejong@utwente.nl)

### Abstract

*This paper focuses on the interaction between test participants and test facilitator in two variants of the think-aloud method. In a first, explorative study, we analyzed think-aloud transcripts from two usability tests: a concurrent think-aloud test and a constructive interaction test. The results of our analysis show that while the participants in both studies never explicitly addressed the facilitator, the think-aloud participants showed more signs of awareness of the facilitator than the participants in the constructive interaction test. This finding may have practical implications for the validity of the two methods.*

**Keywords:** *usability testing, think-aloud protocols, constructive interaction, conversation analysis*

### Introduction

Two of the most popular usability tests are the concurrent think-aloud method and constructive interaction. In both tests, participants evaluate a particular object by verbalizing their thoughts while working with the product, usually guided by a set of fixed tasks. The difference between the two methods is that the concurrent think-aloud method involves only one participant per test session, while constructive interaction involves two participants, who perform their tasks as a team [1].

Previous studies have demonstrated that this difference in number of participants per test session does not necessarily affect the outcome of the two tests with respect to task performance and/or number and types of problems detected [2,

3]. These results seem to favor the use of the concurrent think-aloud method (CTA), as a constructive interaction test (CI) would cost more money (to hire the second participant) to deliver the same results.

Yet the CTA method has some drawbacks of its own. One of these drawbacks concerns the ecological validity of the method. In order to reveal the problems that a user would experience with a product in everyday life, it is vital for a think-aloud study to be conducted in as natural a setting as possible. Nevertheless, CTA tests often take place in test labs, which, while typically cleverly designed, often fail to resemble a natural, everyday environment. Combined with the presence of recording equipment and a test monitor in the test lab, participants in a CTA test are not likely to feel very much at home.

Needless to say, these unnatural conditions also apply to participants in a constructive interaction test. It could be argued, however, that CTA participants are more affected by the presence of the test monitor than CI participants, as they find themselves in a less natural communicative situation. This situation can be ascribed to the standard behavior of test monitors during a concurrent think-aloud test. Very often, the test monitor features in a think-aloud session as a passive, silent witness, who the participant is instructed to ignore completely while performing his tasks and verbalizing his thoughts.

That this standard behavior of the test monitor creates a highly unnatural communicative situation is pointed out by, among others, Boren and Ramey [4]. They discuss one of Bakhtin's seminal works [5] in the context of usability evaluation, paying

particular attention to his claim that there is no such thing as a passive listener. The test monitor, according to Boren and Ramey, may pretend to be a passive listener by having the participant ignore him, but that is not what she is: she necessarily responds to the participants' verbalizations, if only by tacitly approving or disapproving of them, and that gives her an active, albeit silent role in a dialogue between her and the participants. In response, the participants cannot ignore the test monitor as an active listener and will invariably gear their thoughts towards the test monitor. Put simply, even if the participants do not explicitly address the test monitor and the test monitor does not explicitly respond to the participants' verbalizations, there is still an undeniable, implicit dialogue between the two parties.

In a constructive interaction test, the communicative situation is inherently different from that of a CTA test, as it involves three instead of two people: the test monitor and the two participants working together as a team. Several studies have claimed that the presence of these three people allows for a more natural communicative situation: the participants find an active listener in their team mate, which makes it easier for them to forget about the passive listener that is the test monitor [6, 7].

If the above claims are true, then much can be said for preferring the CI test, and not the CTA test, as a method of evaluation. So far, however, none of the claims have been supported by any linguistic evidence of the degree of (un)naturalness in the respective methods. In this paper, we will make a first attempt at addressing this lack of evidence by linguistically analyzing five transcripts from each usability method. Given the limited space available, we will not discuss the entire transcripts, but focus on verbal signposts of participant awareness of the test monitor. These will be illustrated with relevant examples from the transcripts.

### Case study

The transcripts that we analyzed come from two different studies. The first, from which we randomly selected the five CTA transcripts, involved 40 participants: 20 in a concurrent think-aloud condition and 20 in a retrospective think-aloud condition (RTA). [8] The second study, which provided our random selection of CI

transcripts, involved 80 participants: 20 CTA and 20 RTA participants, as well as 20 CI teams. [3]

Both tests involved similar test objects (online library catalogues) and an identical experimental procedure. All participants received the same instructions at the start of their test session. These instructions, which were read from paper so as to ensure consistency, were 'standard' instructions (see the introduction above): they explicitly told the participants to ignore the test monitor who was present during the session. In addition, the CTA participants were instructed to verbalize their *own thoughts* during the task performance, while the CI participants were instructed to work and converse together like a normal team. To ensure that the CTA participants would indeed verbalize their own thoughts and that the CI participants would converse as naturally as possible, we made no attempt to guide the participants' verbalizations: we did, for instance, not instruct the participants to explain their choices during their task performance nor did we tell them that they should refrain from describing what they were doing during their task performance.

### Analysis of the CTA Transcripts

Let us first consider the five transcripts from the CTA study. Analysis of these transcripts with respect to participant awareness of the test monitor revealed two important observations. To begin with, the vast majority of verbalizations in the transcripts consisted of descriptions of actions, such as "I'm clicking on search", "I enter only his surname", etc. (Participant 2, task 2).

This observation is interesting, as the CTA participants were instructed to verbalize their thoughts, not to describe what they were doing. Their behavior would be logical if the participants were indeed thinking about and thus verbalizing their actions. However, as Ericsson and Simon argue [9], actions are typically too automatic to be registered and consequently verbalized as thoughts. This means that the participants must have made a conscious decision to utter them.

One likely explanation for this decision is that the participants felt that they should somehow fill in the gaps that occurred in the absence of thoughts to express. Alternatively, the participants could have felt that it was in the interest of the evaluation to describe their every step. In any case, the CTA

participants' behavior suggests that they were certainly aware of the presence of the test monitor: after all, if they did not acknowledge her as a listener to their verbalizations, they would not have searched for ways to avoid (awkward) silences in their verbalizations or offered complete step-by-step descriptions of their task performance.

The second observation that can be made about the transcripts is that participants were not only aware of the test monitor as listener but also as *evaluator* of their performance. Even though the test monitor had ensured them at the beginning of the test that it was the catalogue and not they that would be evaluated, the participants showed clear signs of feeling tested.

There are several instances in the transcripts which reveal that the participants were eager to perform as best as they could, i.e., in accordance with the proper procedure that they felt they should find and follow. Figure 1 shows some examples of participants who referred to the existence of such a procedure.

... I'm not really sure if I did this correctly or not  
.... (Participant 13, Task 6)

... I think it has to be this, so 38 .... 38, and let's hope it's okay .... (Participant 5, Task 7)

**Figure 1.** Examples of CTA participants referring to the procedure.

What is also illustrated by the examples in Figure 1 is that the participants were clearly insecure about their performance. It seems as if they were almost apologizing to the test monitor for their actions and the possibly unsuccessful result thereof. This apologetic behavior was not uncommon in the five transcripts, and it was often accompanied by the kind of verbalizations illustrated in Figure 2.

... I can't see which (results) are all Dutch. No, I don't think I can find them. Limits then.... No, eh, I can't find it here. (Participant 8, Task 5)

Now I also see that I, eh, maybe should've clicked on 'topic', eh, in the previous task, so that I might've searched more selectively.... (Participant 20, Task 3)

**Figure 2.** Apologetic behavior in CTA participants.

The verbalizations in Figure 2 show that the participants felt that it was up to them to perform their tasks correctly. Rather than pointing out problems with the online catalogue, they blamed themselves for not finding the right results or hitting the correct button. It could be argued that this behavior has no substantial effect on the validity of the CTA method, since the test monitor will know that something is wrong with the site as soon as participants start doubting their skills. However, in blaming incorrect task performances on themselves rather than on the test object, the participants' verbalizations are likely to contain only few linguistic signals pointing to flaws in the catalogue. This can make it quite difficult for test monitors or designers to detect the source of a particular problem, especially when they are testing an object with few visual cues, like a site consisting primarily of plain text.

In sum, this short analysis of the CTA transcripts suggests that the participants in the CTA test were very aware of the test monitor, both as a listener and as an evaluator of their actions. The fact that they were nevertheless supposed to ignore her did create an unnatural communicative situation for them.

### Analysis of the CI Transcripts

Let us now briefly discuss the five transcripts from the constructive interaction test. Our analysis of these transcripts showed that they differed from the CTA transcripts in two respects. First, unlike the CTA participants, the CI participants hardly described any of their actions. As they were working together side by side, they obviously had no reason to describe their actions to their partner. It seems that their working together made them oblivious to the test monitor, in the sense that they did not regard her as a listener who should be kept informed of what was going on at all times.

A second difference between the CTA participants and CI participants is that the latter behaved less insecurely. Figure 3 shows two examples of conversation between team members who blame the site, and not themselves for any unsuccessful actions. Their more confident behavior may be explained by the fact that they worked in a team: they were never alone in not being able to perform a particular action, which gave them more reason to believe that the fault lay not with them, but with the site.

Co-Actor:	I think this text ( <i>in the help section</i> ) is not very clear ....
Actor:	No, you can't tell how truncation works .... (Team 9, Task 5)
Co-Actor:	It's a bit difficult that there's no option to reduce or widen your search ....
Actor:	No, you'd think that'd be possible .... (Team 15, Task1)

**Figure 3.** Example of CI participants blaming the site and not themselves.

For all the difference between CTA and CI participants, the CI participants did resemble the CTA participants in one respect: they too acknowledged the test monitor as an evaluator of their actions. Figure 4 shows two examples of teams who felt that they should follow a proper procedure and arrive at the correct result.

Actor:	Well, this isn't right ....
Co-Actor:	No ... weird ... yeah, we could pick the right ones out, but that's not really how you're supposed to do it .... (Team 17, Task 3)
Co-Actor:	Do you think this is it?
Actor:	Well, do you want to look further? (Team 16, Task 2)

**Figure 4.** CI teams acknowledging that they should follow a proper procedure and arrive at a correct result.

Apparently, the CI participants in the five transcripts were not unaware of the test monitor, even though they benefited from a more natural communicative situation than the CTA participants. What this finding means for the validity of the usability test will be discussed in the conclusion below.

## Conclusion

As this brief analysis has shown, the presence of the test monitor has most notably affected the CTA participants but has also had its impact on the CI participants. These findings may have consequences for the validity of both methods. For the CTA method, the fact that participants in this test verbalized their actions, and hence not their

real thoughts, to keep the test monitor informed of what was happening could lead to the conclusion that they were not contributing towards the detection of real-life user problems, thereby defeating the purpose of the usability evaluation.

Fortunately, however, the success of a think-aloud test does not so much depend on the truthfulness of the participants' expression of their thoughts, as is the case in many psychological experiments, but on the usefulness of these thoughts in detecting user problems. The participants in the CTA test may not have been expressing their actual thoughts while verbalizing their actions, but in verbalizing these actions, they still aided in the detection of user problems.

A more serious threat to the validity of both the CTA method and the CI method is the fact that the participants acknowledged the test monitor as an evaluator of their actions. Their wish to follow a proper procedure made it unlikely that they acted and thought as they would in a real-life situation, which could have caused them to detect different numbers and/or kinds of problems than if they had been working with the test product in a natural environment.

What seems needed, then, is a way to convince the participants that it is really not them, but the test product that is being tested, and, where CTA participants are concerned, that they should not blame themselves for any mistakes or problems that occur during the test session. The instructions to this purpose at the beginning of the tests clearly had no effect, but they might gain more status in future when accompanied by an announcement that the test product really requires some serious testing, for instance because it is not yet up to standard. Knowing that they are working with a not yet perfect version of the test product may encourage the participants to perform a less competitive evaluation. Such non-competitive behavior would greatly enhance the status of the two methods as valid evaluation tools.

## References

- [1] J. Nielsen, *Usability Engineering*, Boston: Academic Press, 1993.
- [2] N. Grondin, J. M. Christian Bastien, and B. Agopian, "Les tests utilisateurs: Avantages et inconvénients des passations individuelles et par



paires”, *Proceedings of the 14th French-speaking conference on Human-computer interaction*, 2002.

[3] M. J. van den Haak, M. D. T. de Jong, and P. J. Schellens, “Employing Think-aloud Protocols and Constructive Interaction to Test the Usability of Online Library Catalogues: A Methodological Comparison”, *Interacting with Computers*, vol. 16, no. 6, pp. 1153-1170, 2004.

[4] T. M. Boren and J. Ramey, “Thinking Aloud: Reconciling Theory and Practice”, *IEEE Transactions on Professional Communication*, vol. 43, no. 3, pp. 261-278, 2000.

[5] M. M. Bakhtin, “The Problem of Speech Genres”, in *Speech Genres and Other Late Essays*, C. Emersen and M. Holmquist, Eds. Austin: University Texas Press, 1986.

[6] J. Höysniemi, P. Hämäläinen, and L. Turkki, “Using Peer Tutoring in Evaluating the Usability of a Physically Interactive Computer Game with Children”, *Interacting with Computers*, vol. 15, pp. 203-225, 2003.

[7] D. Wildman. “Getting the most from paired-user testing”, *ACM Interactions*, vol. 2, no. 3, pp. 21-27, 1995.

[8] M. J. van den Haak, M. D. T. de Jong, and P. J. Schellens, “Retrospective versus Concurrent Think-aloud Protocols: Testing the Usability of an Online Library Catalogue”, *Behaviour and Information Technology*, vol. 22, no. 5, pp. 339-351, 2003.

[9] K.A. Ericsson and H.A. Simon, *Protocol Analysis: Verbal Reports as Data*, Cambridge, MA: MTT Press, 1984.

### About the Authors

Maaïke J. van den Haak is a part-time PhD candidate at the University of Twente (The Netherlands). Her PhD research focuses on the merits and drawbacks of variants of the think-aloud method as an evaluation tool for instructive communication. Apart from her position at the University of Twente, she is also a part-time teacher of English and Translation at the Vrije Universiteit, Amsterdam (The Netherlands).

Menno de Jong is an associate professor of communication studies at the University of Twente, Netherlands. His main research interest concerns the methodology of applied communication research. He has published many articles about document and Web site evaluation and usability testing, and is currently working on an additional research line focusing on applied research methods in organizational and corporate communication.